



TR-6470-2

②
DTIC
S **ELECTE** **D**
MAY 1 1992
C

QUARTERLY TECHNICAL REPORT

15 April 1992

Prepared for:

THE DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

Under:

Contract No. MDA972-92-C-0010

Prepared by:

Karl S. Keller

Approved by:

Amir Eiger

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

TASC
1101 Wilson Boulevard
Suite 1500
Arlington, Virginia 22209

92-11713

92 4 28 3 4 3

INTRODUCTION

This document represents the second quarterly technical report under contract MDA972-92-C-0010 for contract deliverable item 0002AB.

OBJECTIVES

The primary technical objectives of this quarter of the project were to:

- Implement the demonstration prototype including the user interface and the interface to the CBR mechanisms
- Build the demonstration case libraries for the missile domain using available TASC Acquisition data and domain knowledge
- Identify sources for missing program and test data
- Complete the demonstration plan.

TECHNICAL PROBLEMS

The primary technical problem this quarter was development of an interface between our selected user interface package, G2, and the selected CBR prototyping tool, Remind. G2 was to provide the user interface for defining target cases, controlling user queries and activating case retrievals. During this reporting period we discovered that G2 could not support the development of a "strong data type" editing mechanism, and consequently the interface development, without excessive programming. This was surprising given the strongly typed programming interface that resides in G2. We expected to use components of that interface to control the user definition of target cases and questions. Unfortunately, G2's basic implementation required modification to support this goal.

A second technical problem had to do with the appropriateness of the Remind case representation (and corresponding retrieval algorithms) for three of the four case libraries. While the program overview data was primarily symbolic, we discovered that the remaining libraries had mostly textual data for the features of those libraries. Initial data collection did not reveal any symbolic hierarchies that could be constructed to adequately represent the data.

Based on the above problems, the lagging delivery schedules for Remind, and the arrival of the newest version of Inference's CBR Kernel, we chose to continue development of the Phase II prototype using Inference's CBR Kernel. Development of the interface using this product is simplified by an existing MOTIF-based interface tool. Further, the product allows multiple case libraries to be loaded and managed. CBR Kernel is built on top of the schema representation available in ARTIM and employs an object-oriented interface to the CBR functions. Retrieval mechanisms use knowledge of the data type of the features in a case (as in Remind) to employ the correct scoring functions.

GENERAL METHODOLOGY

The primary methodology for implementing the prototype has not changed since the last quarter. Rapid prototyping is being used to develop interface components and their sub-functions. The case bases used during the initial prototyping effort in Remind were translated directly to the ARTIM representations.

When possible, data is imported electronically into the system to avoid transcription errors. This has worked effectively for all of the TASC Acquisition Data and will be the mode of operation for data collected from outside sources as well.

TECHNICAL RESULTS

The primary technical results of the quarter relate to the implementation of the demonstration prototype. Several development prongs were initiated and completed during this quarter:

1. The development of a nearest neighbors algorithm (for symbol hierarchies) in ARTIM to augment the text-based mechanisms in CBR Kernel
2. The development of a natural language parser to support the query generation mechanism of the interface
3. The development of a strongly data typed editing mechanism for describing target cases or augmenting the case libraries

Accession For	
NTIS GR&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By: AD-A248268	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	



4. The development of a MOTIF-based user interface to display the case data (both graphically and textually) and for activating the retrieval and case comparison process
5. The completion of the demonstration plan
6. The development of a fifth case library for storing the recommendations of previous program managers (to complete the importing of the TASC Acquisition database).

In addition to the technical results listed, we have undertaken to identify data resources to be used during the upcoming quarters to begin the collection and population of missile program test cases.

IMPORTANT FINDINGS OR CONCLUSIONS

The most important discovery of this quarter relates to the symbolic vs. textual case data problems noted in the opening paragraphs. The derivation of symbolic concept hierarchies from free text (actually, short segments of free text) is an interesting machine learning research question. For our purposes, collection of a larger case library (with a greater number of textual examples) could begin to uncover some emerging structure in the features of interest. As the system development progresses, certain features could be migrated from the textual representation to a symbolic representation. A mechanism for doing this automatically would be very useful.